Semi-Markov Arnason-Schwarz models

Biometrics ??, 1–24 DOI: 10.1111/???
XXXXX XXXX

Semi-Markov Arnason-Schwarz Models

Ruth King[1,*] and Roland Langrock[2,]**

[1]School of Mathematics, University of Edinburgh, James Clerk Maxwell Building,
The King's Buildings, Peter Guthrie Tait Road, Edinburgh, UK, EH9 3FD.

[2]Department of Business Administration and Economics,
Bielefeld University, Postfach 10 01 31, 33501 Bielefeld, Germany.

email: Ruth.King@ed.ac.uk

**email:* roland.langrock@uni-bielefeld.de

Summary: We consider multi-state capture-recapture-recovery data where observed individuals are recorded in a set of possible discrete states. Traditionally, the Arnason-Schwarz model has been fitted to such data where the state process is modeled as a first-order Markov chain, though second-order models have also been proposed and fitted to data. However, low-order Markov models may not accurately represent the underlying biology. For example, specifying a (time-independent) first-order Markov process involves the assumption that the dwell time in each state (i.e., the duration of a stay in a given state) has a geometric distribution, and hence that the modal dwell time is one. Specifying time-dependent or higher-order processes provides additional flexibility, but at the expense of a potentially significant number of additional model parameters. We extend the Arnason-Schwarz model by specifying a semi-Markov model for the state process, where the dwell-time distribution is specified more generally, using for example a shifted Poisson or negative binomial distribution. A state expansion technique is applied in order to represent the resulting semi-Markov Arnason-Schwarz model in terms of a simpler and computationally tractable hidden Markov model. Semi-Markov Arnason-Schwarz models come with only a very modest increase in the number of parameters, yet permit a significantly more flexible state process. Model selection can be performed using standard procedures, and in particular via the use of information criteria. The semi-Markov approach allows for important biological inference to be drawn on the underlying state process, for example on the times spent in the different states. The feasibility of the approach is demonstrated in a simulation study, before being applied to real data corresponding to house finches where the states correspond to the presence or absence of conjunctivitis.

Key words: capture-recapture-recovery; dwell-time distribution; hidden Markov model; multi-state model.

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

Capture-recapture studies, for example on wildlife populations, involve researchers surveying a given population at a series of capture events. At the initial capture event all individuals observed are uniquely marked, recorded and released. At each subsequent capture event, all individuals observed are recorded, marked if they have not previously been observed and again released. The additional observation of dead individuals leads to capture-recapture-recovery data. Such studies have been formulated within a state-space or hidden Markov model (HMM) framework, for example, in Gimenez *et al.* (2007), Royle (2008) and King (2012, 2014), which has proven particularly useful for scenarios which additionally involve individual-level covariate data. We consider the case of time-varying discrete individual covariates (Lebreton *et al.*, 2009). Such covariates may relate to breeding status, disease status, foraging/resting, life stage, mark status, etc. The covariate values are often referred to as 'states' and we use these terms interchangeably. Typically it is assumed that if an individual is observed, their corresponding state is also observed. The observed data are the capture/encounter histories for each individual observed within the study. For example, assuming a single discrete covariate, an observed capture history for an individual may be:

$$3 \quad 0 \quad 0 \quad 2 \quad 3 \quad 0 \quad 0 \quad 0 \quad \dagger$$

This history represents an individual that is observed at the initial capture time in state 3, recaptured at times 4 and 5, in states 2 and 3, respectively, and recovered dead at the final capture occasion. A '0' indicates that an individual is unobserved at the given capture time.

The Arnason-Schwarz (AS) model is often fitted to multi-state capture-recapture(-recovery) data (Brownie *et al.*, 1993; Schwarz *et al.*, 1993; King and Brooks, 2002, 2003; for a review see Lebreton *et al.*, 2009). The model typically assumes the state process is a first-order Markov chain. Under time-homogeneity, this means that the length of time an individual remains in a state is independent of the length of time the individual has already spent in

the state (i.e., the process is memoryless). Further, the times spent in the states (i.e., the dwell times) are geometrically distributed with the most likely duration of a stay in a state one time unit (conditional on being alive). These assumptions facilitate the implementation, but are often not biologically realistic. In particular, the probability that an individual leaves a state is often a function of the length of time the individual has been in the state. For example, consider the two states corresponding to 'exploratory/foraging' and 'encamped' – an individual will typically forage until its hunger is satiated which in general will be a function of the length of time it has been foraging (see, for example Langrock *et al.*, 2012, for further discussion). Alternatively state may correspond to the disease states 'susceptible' and 'infected' with the length of time an individual is infected having a non-geometric distribution – the disease follows a biological process where the recovery time may follow a distribution with a mode distinct from one. In general, the dwell time distributions of the states may themselves be of biological interest (for example where state refers to life-stage). Higher-order Markov processes can be specified: for example, Hestbeck *et al.* (1991), Brownie *et al.* (1993) and Rouan *et al.* (2009) consider a second-order state process. However, the number of transition parameters increases exponentially as the order increases, so that consideration of processes of order greater than two is usually infeasible, and such models provide only limited additional memory structure. Semi-Markov models constitute a different, parsimonious class of models where the dwell-time distribution of each state is explicitly modelled. These models can be interpreted as age-dependent models, where 'age' corresponds to length of time spent in state. This idea has been applied to stopover models (Pledger *et al.*, 2009; Fewster and Patenaude, 2009; Choquet *et al.*, 2013) where the length of time that an individual is present at a site is dependent on their arrival time, and to multi-state capture-recapture models under certain constraints (Choquet *et al.*, 2014). In particular, Choquet *et al.* (2014) consider a semi-Markov multi-state capture-recapture model, but focus on a single semi-Markovian state

where the states are ordered so that individuals can only enter the semi-Markov state once. Their model-fitting approach, specifying the likelihood by summing over all entry and exit times to the semi-Markovian state, becomes infeasible for multiple or unordered processes.

We describe the general class of semi-Markov AS models, where a distribution on the positive integers – for example, a Poisson shifted by one (with support $1, 2, \ldots$), a negative binomial shifted by one, simply the geometric distribution (in which case the state process is Markovian) or an arbitrary discrete distribution on the positive integers – is specified on the dwell time for each state. Together with the conditional state transition probabilities, given an individual being alive and a state being left, this determines the semi-Markov model for the covariate process. Further, we describe an efficient model-fitting approach for the proposed class of semi-Markov AS models. The semi-Markov AS model specification is substantially more flexible than a first-order Markov model, yet has only very few (if any) additional parameters. In addition, it is possible to apply standard model selection techniques (for example, information criteria or likelihood ratio tests) and test for specific biological hypotheses such as the absence or presence of memory in the different states (i.e., comparing first-order memoryless Markov models with semi-Markov memory models).

The manuscript is structured as follows. Section 2 gives the model formulation and associated estimation methods. A simulation study is provided in Section 3, before we analyze capture-recapture data of house finches, where state corresponds to presence/absence of conjuctivitis, in Section 4. Section 5 concludes with final remarks.

2. Semi-Markov AS model

2.1 *Mathematical notation*

2.1.1 *Observations.* We let N denote the number of individuals observed within the study, T the total number of capture occasions and $\mathcal{K} = \{1, \ldots, K\}$ the set of possible discrete

covariate values. For each individual $i = 1, \ldots, N$, and capture time $t = 1, \ldots, T$,

$$x_{it} = \begin{cases} 0 & \text{if individual } i \text{ is not observed at time } t; \\ k & \text{if individual } i \text{ is observed alive in state } k \in \mathcal{K} \text{ at time } t; \text{ and} \\ \dagger & \text{if individual } i \text{ is recovered dead in the interval } (t-1, t]. \end{cases}$$

Letting t_{i0} denote the initial capture time for individual i, we define $\mathcal{W}_i = \{t \geq t_{i0} : x_{it} \in \{1, \ldots, K, \dagger\}\}$ as the set of times that individual i is observed, and let $\mathcal{W}_i^c$ denote its complement as the set of times the individual is not observed but potentially alive following their initial capture. We assume that covariate values are observed without error. This assumption will be relaxed in Section 4.2 where covariates may be unknown when an individual is observed.

2.1.2 *Survival and covariate states.* We now consider the underlying state process and the associated notation. For individual $i = 1, \ldots, N$, we let s_{it} denote the state of the individual at time t, such that, for $t = t_{i0}, \ldots, T$:

$$s_{it} = \begin{cases} k & \text{if individual } i \text{ is alive and has discrete covariate value } k \in \mathcal{K} \text{ at time } t; \\ K+1 & \text{if individual } i \text{ is dead at time } t, \text{ but was alive at time } t-1; \\ K+2 & \text{if individual } i \text{ is dead at time } t, \text{ and was dead at time } t-1. \end{cases}$$

We distinguish "recently dead" (state $K+1$) from "long dead" (state $K+2$) individuals, since it is typically assumed that individuals can only be recovered dead in the same interval that death occurs. Model adaptations removing this assumption are straightforward and discussed in Section 2.2.2. For notational convenience we drop the subscript i corresponding to individual, but make it clear in the text that we are referring to the individual.

2.2 *Traditional AS model*

2.2.1 *Model parameters.* We initially consider the standard multi-state AS model, which assumes a first-order Markov state process. We allow for both live recaptures and dead recoveries. For $j, k \in \mathcal{K}$, the model parameters are defined as follows:

$$\begin{aligned}
\phi_t(k) &= \Pr(s_{t+1} \in \mathcal{K} | s_t = k) && (\textit{survival probabilities});\\
\psi_t(j,k) &= \Pr(s_{t+1} = k | s_t = j \text{ and } s_{t+1} \in \mathcal{K}) && (\textit{transition probabilities});\\
p_t(k) &= \Pr(x_t = k | s_t = k) && (\textit{recapture probabilities});\\
\lambda_t &= \Pr(x_t = \dagger | s_t = K+1) && (\textit{recovery probabilities});\\
\pi_{t_0}(k) &= \Pr(s_{t_0} = k) && (\textit{initial state probabilities}).
\end{aligned}$$

2.2.2 *Formulation as HMM.* We extend the notation and HMM-type specification of the Cormack-Jolly-Seber (CJS) model for capture-recapture-recovery data presented by Langrock and King (2013). The model is specified as a partially observed state process, coupled with an observation process, defined conditional on the underlying state. The state process is essentially decomposed into the survival process (i.e., whether or not an individual remains in $\mathcal{K}$) and the transitioning process between covariate states, conditional on survival. The state process has probability mass function (p.m.f.)

$$f(s_{t+1}|s_t) = \begin{cases}
\phi_t(s_t)\psi_t(s_t, s_{t+1}) & \text{for } s_t, s_{t+1} \in \mathcal{K};\\
1 - \phi_t(s_t) & \text{for } s_t \in \mathcal{K},\ s_{t+1} = K+1;\\
1 & \text{for } s_t \in \{K+1, K+2\},\ s_{t+1} = K+2;\\
0 & \text{otherwise.}
\end{cases}$$

It is immediate to show that $\sum_{k=1}^{K+2} f(k|s_t) = 1$ for $s_t = 1, \ldots, K+2$.

The p.m.f. for the observation process is defined as follows:

$$f(x_t|s_t) = \begin{cases}
p_t(s_t) & \text{for } x_t = s_t,\ s_t \in \mathcal{K};\\
1 - p_t(s_t) & \text{for } x_t = 0,\ s_t \in \mathcal{K};\\
\lambda_t & \text{for } x_t = \dagger,\ s_t = K+1;\\
1 - \lambda_t & \text{for } x_t = 0,\ s_t = K+1;\\
1 & \text{for } x_t = 0,\ s_t = K+2;\\
0 & \text{otherwise.}
\end{cases}$$

It is straightforward to modify this definition to deal with scenarios in which "long dead"

individuals can be recovered within the study (see for example Catchpole *et al.*, 2001). For such cases, the recovery probability may be constant for all dead individuals or specified as a decreasing function of time following death.

2.2.3 *Likelihood.* The likelihood is the product over each individual of the corresponding probability of their observed encounter history, conditional on their initial capture. The likelihood component for an individual initially observed at time t_0 can be calculated by summing over all possible unknown states after initial capture. Mathematically,

$$\mathcal{L} = \sum_{\tau \in \mathcal{W}^c} \sum_{s_\tau \in \{1,\ldots,K+2\}} \pi(s_{t_0}) \prod_{t=t_0}^{T-1} f(s_{t+1}|s_t) f(x_{t+1}|s_{t+1}).$$

This likelihood can be conveniently and computationally efficiently calculated using an HMM-type forward algorithm. In order to see this, we define the time-dependent transition probability matrix (t.p.m.) for the states as

$$\mathbf{\Gamma}_t(x_t, x_{t+1}) = \begin{pmatrix} \phi_t(1)\tilde{\psi}_t(1,1) & \cdots & \phi_t(1)\tilde{\psi}_t(1,K) & 1-\phi_t(1) & 0 \\ \vdots & & \vdots & \vdots & \vdots \\ \phi_t(K)\tilde{\psi}_t(K,1) & \cdots & \phi_t(K)\tilde{\psi}_t(K,K) & 1-\phi_t(K) & 0 \\ 0 & & \cdots & 0 & 1 \\ 0 & & \cdots & 0 & 1 \end{pmatrix},$$

where

$$\tilde{\psi}_t(j,k) = \begin{cases} \psi_t(j,k) & \text{for } (x_t, x_{t+1}) \in \{(j,k),(0,k),(j,0),(0,0)\}; \\ 0 & \text{otherwise,} \end{cases}$$

for $j, k \in \mathcal{K}$. Note that when an individual is observed and the corresponding covariate value recorded at successive times t and $t+1$, the associated $\mathbf{\Gamma}_t$ matrix reduces such that there is only a single non-zero element in the $(K \times K)$ upper-left submatrix. Similarly if the covariate value is observed at time t and not $t+1$, the upper-left submatrix of $\mathbf{\Gamma}_t$ is composed of a single row of non-zero elements; if the covariate value is observed at time $t+1$ and not time t there is a single column of non-zero elements.

The state-dependent observation process is expressed as a diagonal matrix of the form:

$$\boldsymbol{Q}_t(x_t) = \begin{cases} \text{diag}(1-p_t(1),\ldots,1-p_t(K),1-\lambda_t,1) & \text{if } x_t = 0; \\ \text{diag}(0,\ldots,p_t(k),\ldots,0,0,0) & \text{if } x_t = k \in \mathcal{K}; \\ \text{diag}(0,\ldots,0,\lambda_t,0) & \text{if } x_t = \dagger. \end{cases}$$

Putting all the pieces together, the corresponding likelihood contribution for the given individual, conditional on being initially observed at time $t = t_0$, is:

$$\mathcal{L} = \boldsymbol{\pi}_{t_0} \left(\prod_{t=t_0}^{T-1} \boldsymbol{\Gamma}_t(x_t, x_{t+1}) \boldsymbol{Q}_{t+1}(x_{t+1}) \right) \mathbf{1}_{K+2},$$

where $\mathbf{1}_{K+2}$ denotes a column vector of ones of length $K+2$, and $\boldsymbol{\pi}_{t_0}$ the initial state distribution at the initial capture, given that the individual is observed. It will often be reasonable to assume that the initial detection does not depend on state, in which case for the time-homogeneous case (for transitions between covariate states) the initial state probability can conveniently be taken as the stationary distribution of the state process conditional on the animal being alive, i.e., the solution to

$$\tilde{\boldsymbol{\pi}}_{t_0} = \tilde{\boldsymbol{\pi}}_{t_0} \begin{pmatrix} \psi(1,1) & \cdots & \psi(1,K) \\ \vdots & & \vdots \\ \psi(K,1) & \cdots & \psi(K,K) \end{pmatrix}.$$

Under this assumption, for an individual observed in state $k \in \mathcal{K}$ at initial capture, $\boldsymbol{\pi}_{t_0}$ is the vector of length $K+2$ with k-th entry $\tilde{\pi}_{t_0}(k)$ and all other entries 0. Alternatively, and more generally, the initial state distribution can be estimated (see for example Pradel, 2009). Finally, if we condition on the initial observed covariate value, $\boldsymbol{\pi}_{t_0}$ is the vector with the kth entry equal to one and all other entries 0.

2.3 *Semi-Markov AS model*

2.3.1 *Non-geometric dwell-time distributions.* We extend the AS model to allow for semi-Markov state processes, relaxing the restrictive condition that, while alive, the dwell-time distribution in a state follows a geometric distribution. The observation process of a semi-

Markov AS model and the survival, recapture and recovery probabilities are defined as for the basic AS model. However, in a semi-Markov AS model, the dwell-time in each state, conditional on survival, can follow any discrete distribution on the natural numbers, for example, a shifted Poisson or shifted negative binomial. The state process is determined by the survival probabilities, the p.m.f.s of the state dwell-time distributions, $d_1(r), \ldots, d_K(r)$, $r = 1, 2, \ldots$ – with associated cumulative distribution functions $F_1(r), \ldots, F_K(r)$ – and the conditional state transition probabilities, given the current state is left:

$$\psi_t^*(j,k) = \mathbb{P}\big(s_{t+1} = k | s_t = j \text{ and } s_{t+1} \in \mathcal{K} \setminus \{j\}\big),$$

for $j \neq k$. The probabilities of self-transitions, $\psi_t(k,k)$, conditional on survival, are determined by the dwell-time distributions, $d_k(r)$. Consequently, the state process is no longer Markovian, with the probability an individual leaves a state dependent on how long they have been in the state. However, we note that the embedded sequence comprising of simply the order of the states visited – for example, the subsequence 1, 2, 1, 3 of the sequence 1, 2, 2, 2, 1, 3, 3 – does satisfy the Markov property, and is described by the transition probabilities $\psi_t^*(j,k)$. First-order Markov AS models (with time-homogeneous state transition probabilities) are a special case of semi-Markov AS models where all dwell-time distributions are geometric.

A common distribution on the positive integers is the shifted negative binomial distribution, which we denote by $\text{sNB}(\nu, \theta)$. The associated p.m.f. is given by

$$p(r) = \binom{r+\nu-2}{r-1} \theta^\nu (1-\theta)^{r-1}, \qquad r = 1, 2, 3, \ldots .$$

The (shifted) negative binomial distribution represents a generalization of the geometric distribution: $p(r)$ gives the probability that $r-1$ 'failures' occur before ν 'successes' have occurred ($\nu = 1$ corresponds to the geometric distribution). Using the gamma function, the definition is easily extended to real-valued ν. Other potentially useful dwell-time distributions

include the shifted binomial and the shifted Poisson. More flexible specifications, for example, using mixtures of Poisson distributions, or even such where one parameter is estimated for every point in the support, are equally easy to implement in the framework we propose.

2.3.2 *The likelihood.* The likelihood for models with hidden semi-Markovian state processes is more complex than for the Markovian case. Choquet *et al.* (2014) consider in detail the special case where there is only a single semi-Markovian state, which cannot be revisited once an individual leaves the state (i.e. states are ordered). In this case an explicit expression for the likelihood can be derived by summing over all possible entry and exit times from the given semi-Markovian state. However, the necessary summation becomes infeasible for multiple semi-Markovian states or for multiple visits to a semi-Markovian state, and Choquet *et al.* (2014) suggest the development of an EM algorithm for these more general models.

We propose an alternative general model-fitting algorithm that extends the approach presented in Langrock and Zucchini (2011) to deal with capture-recapture data. More specifically, we construct a specially structured AS model with first-order state process so that it mirrors the properties of the semi-Markov AS model to be fitted. The corresponding model formulation can be used to fit semi-Markov AS models with *any* desired dwell-time distributions. The idea is to use so-called state aggregates, where each of the semi-Markovian covariate states is expanded into a (large) set of Markovian states, structuring the transitions between the latter so that the dwell-time distribution within each state aggregate of the Markov model matches the corresponding dwell-time distribution of the semi-Markov model to be represented. Thus, we represent an arbitrary semi-Markov AS model (i.e. a hidden semi-Markov model) as a simple AS model (i.e. an HMM) with an expanded state space. The latter model is not of interest in itself, but is only used as a tool for fitting the former, semi-Markov model. The advantage of this state expansion is that we can use the HMM representation for conducting statistical inference for the given semi-Markov AS model, which means that the

entire HMM machinery – most notably the forward algorithm for evaluating the likelihood – becomes applicable to the semi-Markov AS model, without any further amendments being required. Langrock *et al.* (2012) explain the state expansion technique for translating a semi-Markov model into a Markov model by means of a simple example. The general model formulation is presented in the following.

Let $a_1, a_2, \ldots, a_K$ be positive integers, and let $a_0 = 0$. We consider a first-order Markov chain with $a^* = \sum_{l=1}^{K} a_l + 2$ states, where state $k \in \mathcal{K}$ of the semi-Markov AS model is expanded into a set of a_k states. We refer to the sets $I_k = \left\{ n \mid \sum_{l=0}^{k-1} a_l < n \leq \sum_{l=0}^{k} a_l \right\}$ as *state aggregates*, and assume that each state of I_k is associated with the same distribution for the observation process, namely that associated with state k of the semi-Markov AS model. For $k = 1, \ldots, K$ and $r = 1, 2, \ldots$, let $c_k(r) = \{d_k(r)\}/\{1 - F_k(r-1)\}$ for $F_k(r-1) < 1$, and $c_k(r) = 1$ for $F_k(r-1) = 1$. The functions c_k play a key role in the following; essentially they are responsible for rendering the desired dwell-time distributions. The $a^* \times a^*$ t.p.m. of the model with first-order state process that represents the semi-Markov AS model is defined as:

$$\mathbf{\Gamma}_t^*(x_t, x_{t+1}) = \begin{pmatrix} \phi_t(1)\mathbf{\Gamma}_{11,t}^* & \cdots & \phi_t(1)\mathbf{\Gamma}_{1K,t}^* & (1-\phi_t(1))\mathbf{1}_{a_1} & 0 \\ \vdots & & \vdots & \vdots & \vdots \\ \phi_t(K)\mathbf{\Gamma}_{K1,t}^* & \cdots & \phi_t(K)\mathbf{\Gamma}_{KK,t}^* & (1-\phi_t(K))\mathbf{1}_{a_K} & 0 \\ 0 & & \cdots & 0 & 1 \\ 0 & & \cdots & 0 & 1 \end{pmatrix}. \tag{1}$$

Here, for $(x_t, x_{t+1}) \in \{(k,k), (0,k), (k,0), (0,0)\}$, the $a_k \times a_k$ *leading diagonal matrix* $\mathbf{\Gamma}_{kk,t}^*$, $k = 1, \ldots, K$, is defined as

$$\mathbf{\Gamma}_{kk,t}^* = \left(\begin{array}{c|cccc} 0 & 1-c_k(1) & 0 & \ldots & 0 \\ \vdots & 0 & \ddots & & \vdots \\ & \vdots & & & 0 \\ 0 & 0 & \ldots & 0 & 1-c_k(a_k-1) \\ \hline 0 & 0 & \ldots & 0 & 1-c_k(a_k) \end{array}\right); \tag{2}$$

else it is the $a_k \times a_k$ matrix with all entries equal to zero. In the special (geometric) case with $a_k = 1$, we define $\mathbf{\Gamma}^*_{kk,t} = 1 - c_k(1)$ for $(x_t, x_{t+1}) \in \{(k,k),(0,k),(k,0),(0,0)\}$, and $\mathbf{\Gamma}^*_{kk,t} = 0$ otherwise. (Note that in this case we can drop the notational dependence on t for $\mathbf{\Gamma}^*_{kk,t}$). Furthermore, for $(x_t, x_{t+1}) \in \{(j,k),(0,k),(j,0),(0,0)\}$, the $a_j \times a_k$ *off-diagonal matrix* $\mathbf{\Gamma}^*_{jk,t}$, $j,k = 1,\ldots,K$, $j \neq k$, $t = 1,\ldots,T-1$ is defined as

$$\mathbf{\Gamma}^*_{jk,t} = \begin{pmatrix} \psi^*_t(j,k)c_j(1) & 0 & \ldots & 0 \\ \psi^*_t(j,k)c_j(2) & 0 & \ldots & 0 \\ \vdots & & & \\ \psi^*_t(j,k)c_j(a_j) & 0 & \ldots & 0 \end{pmatrix}. \tag{3}$$

In case of $a_k = 1$, the zeros disappear. Thus, different dwell-time distributions lead to different c_k's, while the $\psi^*_t(j,k)$'s and the structure of the t.p.m. $\mathbf{\Gamma}^*_t(x_t, x_{t+1})$ remain unaffected. The structure of $\mathbf{\Gamma}^*_t(x_t, x_{t+1})$ can be interpreted as follows. Conditional on survival, all transitions within state aggregate I_k are governed by the diagonal matrix $\mathbf{\Gamma}^*_{kk,t}$, which thus determines the dwell-time distribution for I_k. The off-diagonal matrices determine the probabilities of transitions between different state aggregates. For example, for $j \neq k$, the matrix $\mathbf{\Gamma}^*_{jk,t}$ contains the probabilities of all possible transitions between the state aggregates I_j and I_k (between times t and $t+1$). Within this model specification these conditional state transition probabilities, given a state is left, are independent of the length of time spent in a state.

We now consider the initial state distribution, $\boldsymbol{\pi}^*_{t_0}$. Using the first-order representation, it is straightforward to fit a semi-Markov AS assuming that the initial capture is not dependent on state, the covariate state transition process is independent of time and the state process conditional on the animal being alive, i.e., the process restricted to the states $\{1,\ldots,K\}$, is in equilibrium at the initial capture time, t_0. To achieve this, let $\tilde{\boldsymbol{\pi}}^*_{t_0} = (\tilde{\pi}^*_{t_0}(1),\ldots,\tilde{\pi}^*_{t_0}(\sum_{l=1}^K a_l))$

be the solution to the equation system

$$\tilde{\boldsymbol{\pi}}^*_{t_0} = \tilde{\boldsymbol{\pi}}^*_{t_0} \begin{pmatrix} \boldsymbol{\Gamma}^*_{11} & \cdots & \boldsymbol{\Gamma}^*_{1K} \\ \vdots & & \vdots \\ \boldsymbol{\Gamma}^*_{K1} & \cdots & \boldsymbol{\Gamma}^*_{KK} \end{pmatrix},$$

dropping the dependence of $\boldsymbol{\Gamma}_{jk}$ for $j \neq k$ on time, t, under the assumption of time-homogeneity for the covariate state transition process. Assuming stationarity of the restricted state process, if the observed covariate is $k = 1, \ldots, K$ at time t_0, then $\boldsymbol{\pi}^*_{t_0}$ is the vector with entries $l \in I_k$ given by $\tilde{\pi}^*_{t_0}(l)$ and all other entries equal to zero. Alternatively, the initial state distribution can be estimated, in which case, equilibrium within state aggregates may again be assumed for numerical stability. Finally, if we condition on the initial observed state, we would also typically assume stationarity within state aggregates.

We now consider the likelihood

$$\mathcal{L} = \boldsymbol{\pi}^*_{t_0} \left(\prod_{t=t_0}^{T-1} \boldsymbol{\Gamma}^*_t(x_t, x_{t+1}) \boldsymbol{Q}^*_{t+1}(x_{t+1}) \right) \mathbf{1}_{a^*}. \tag{4}$$

The diagonal matrix $\boldsymbol{Q}^*_t(x_t)$ is specified analogously to the definition of $\boldsymbol{Q}_t(x_t)$ in Section 2.2.3, with diagonal entry $k \in \mathcal{K}$ in $\boldsymbol{Q}_t(x_t)$ repeated a_k times on the diagonal in $\boldsymbol{Q}^*_t(x_t)$, corresponding to the expansion of state k to the state aggregate I_k. This representation in general only approximates the semi-Markov model (Langrock and Zucchini, 2011). The p.m.f. of the distribution of the time spent in state aggregate I_k, denoted by $d^*_k(r)$, in general differs from $d_k(r)$ for $r > a_k$, i.e., in the right tail, since by its definition $d^*_k(r)$ exhibits a geometric tail. The difference between d_k and d^*_k can be made arbitrarily small by choosing a_k sufficiently large. Initially, the a_k are chosen a priori, i.e. before fitting the model and hence not knowing the shape of the dwell-time distributions. Whether or not the values chosen for the a_k were indeed sufficiently large should always be verified with a post hoc inspection of the estimated dwell-time distributions – if these have wider supports than expected, then the model might need to be re-fitted using larger values for the a_k. Values

of a_k that are significantly higher than necessary – for example $a_k = 200$ when a Poisson(3) dwell-time distribution is expected for state k – should be avoided due to the associated higher computational effort. For any dwell-time distribution with either finite support or geometric tail, we can ensure that $d_k^*(r) = d_k(r)$ for all r by choosing a_k appropriately. In summary, the likelihood of the semi-Markov AS model can be approximated by (4), where the approximation can be made arbitrarily accurate by choosing sufficiently large values a_k for $k = 1, \ldots, K$.

2.4 *Inference*

For both the standard AS models and the semi-Markov AS models, for multiple individuals, the likelihood is simply the product of likelihoods of type (4), corresponding to each observed capture history. It is then a routine matter to numerically maximize this joint likelihood with respect to the model parameters, subject to well-known technical issues arising in all optimisation problems (for example, parameter constraints, numerical underflow, local maxima). Approximate confidence intervals for the parameters can be obtained based on the inverse of the estimated information matrix, or, alternatively, using a parametric/non-parametric bootstrap. Model selection, including for the underlying covariate process model, can easily be carried out using model selection criteria such as Akaike Information Criterion (AIC).

3. Simulation study

A simulation study was conducted to demonstrate the feasibility of the semi-Markov approach and to investigate potential consequences of fitting simpler Markovian models when the operating model is semi-Markov. We consider a semi-Markov AS model with three underlying covariate states, each with a different dwell-time distribution, namely a shifted negative binomial (with parameters $\nu = 4$ and $\theta = 0.4$), a shifted Poisson (with mean $\mu = 4$) and a geometric (with parameter $\theta = 0.4$), respectively. We assume that the other model

parameters are constant over time, and so omit the t subscript for each of the parameters. The conditional state transition probabilities (given the covariate state is left) are specified as $\psi^*(1,2) = 0.6 = 1 - \psi^*(1,3)$, $\psi^*(2,1) = 0.8 = 1 - \psi^*(2,3)$ and $\psi^*(3,1) = 0.5 = \psi^*(3,2)$. The (state-dependent) survival probabilities are specified as $\phi(1) = 0.8$, $\phi(2) = 0.9$ and $\phi(3) = 0.6$, and recapture probabilities are specified as $p(1) = 0.2$, $p(2) = 0.1$ and $p(3) = 0.5$. The (state-independent) recovery probability is taken to be $\lambda = 0.2$. Sample `R` code for simulating and fitting such a semi-Markov AS model is provided in the Web Appendix.

We simulated 1000 datasets, each with $N = 500$ individuals and $T = 20$ capture events. For each dataset, we fitted (i) the correctly specified semi-Markov AS model and (ii) the corresponding incorrect first-order Markov model. The latter was done in order to investigate potential consequences of not accounting for a semi-Markov structure in the covariate process. Figure 1 displays the p.m.f.s of the three dwell-time distributions, together with the estimated distributions obtained in the first 10 simulation runs when using either the first-order or semi-Markov model specification (these are representative of the results obtained for all the simulated datasets). Table 1 summarizes the results of the remaining parameter estimates obtained using the two different model formulations, by providing the mean relative biases and mean standard deviations.

[Table 1 about here.]

[Figure 1 about here.]

As expected, the dwell-time distribution is inaccurately estimated when fitting the incorrect geometric distribution (i.e., the first-order Markov model) for the two states with non-geometric distributions. We note that, due to its higher flexibility, the uncertainty in the estimation of the negative binomial state dwell-time distribution (in state 1) was found to be much higher than for the dwell-time distributions in the other two states (cf. the top row of plots in Figure 1). In general, the dwell-time distributions of the states may be of interest

themselves, for example, where state refers to life stage which in turn may be used to infer future population trajectories.

The semi-Markov approach led to approximately unbiased estimates for all the parameters except the recapture probabilities (see Table 1). These displayed a positive bias as a result of small-sample properties of the estimation: all mean relative biases decreased to values < 0.005 when sample size was increased by a factor of 10. We note that the recapture parameters are essentially nuisance parameters, i.e., not of primary interest. In contrast, fitting the first-order model led to biased estimates for the conditional state transition probabilities, despite the fact that the embedded subsequence corresponding to the order the states occur in (ignoring duration in each state) is first-order Markovian. Clearly, the inaccurate estimation of the dwell-time distribution and biased estimation of the conditional state probabilities can be problematic when interest lies in the state-switching dynamics. However, interestingly, the survival probabilities were generally robust with regard to the mis-specification of the dwell-time distributions, at least for the set of parameter values used in this simulation study.

4. Application to *Carpodacus mexicanus*

4.1 *Data*

We consider multi-state capture-recapture histories for house finches (*Carpodacus mexicanus*) collected at a study site in Ithaca, New York, between September and December 2002. There are no recoveries in this study. During the study period a total of $N = 813$ individuals are observed over the $T = 16$ capture occasions. The covariate states correspond to the presence/absence of *Mycoplasma gallisepticum* conjunctivitis, so that $K = 2$. We let state 1 denote the absence of conjunctivitis and state 2 the presence of conjunctivitis. For these data, it was not always possible to determine the presence/absence of conjunctivitis (i.e., the state) of an individual bird observed at a given capture occasion: the state is unknown for 4.1% of

the recorded observations. The data are provided in the Web Appendix. Particular interest lies in (i) the survival probabilities for the two states and (ii) the dwell-time distributions for the two states. For further details of the data collection process and the biological mechanisms, see, for example, Faustino *et al.* (2004) and Conn and Cooch (2009).

4.2 *Semi-Markov AS model with unobserved states*

We extend the semi-Markov AS model discussed above to account for the possibility of the covariate state of an individual, here presence/absence of conjunctivitis, to be unknown when an individual is observed. We define $\alpha_t(k)$ to be the assignment probability that the state of an individual is recorded, given that the individual is observed at time t and is in state k (i.e., $s_t = k$). For the special case where all covariates are always observed (without error), the corresponding p.m.f. is trivial (a point mass on the true values, so that $\alpha_t(k) = 1$ for all $k \in \mathcal{K}$). Alternatively, if the covariate values are missing at random, then $\alpha_t(k) = \alpha_t$ for all $k \in \mathcal{K}$, and simply corresponds to the probability that the covariate value is recorded independently of state, given the individual is observed. Finally, if the covariate values are *not* missing at random, then in general $\alpha_t(j) \neq \alpha_t(k)$ for $j \neq k$. The corresponding likelihood in the presence of unobserved states is provided in the Web Appendix for both the standard AS model and the semi-Markov AS model. The extension to partially observed covariate values, or covariate values observed with error (i.e., false reporting of covariate values) can be similarly incorporated but is omitted here to avoid additional complexity (see King and McCrea, 2014, for further discussion of such models).

4.3 *Models considered*

A first-order Markov AS model, allowing for unknown state when an individual is observed, has been fitted to these data by Conn and Cooch (2009), Schofield and Barker (2011) and King and McCrea (2014). In particular, King and McCrea (2014) fit a range of models to

the data and identify the following model as optimal via the AIC statistic:

$\phi(t+c)$: time (t) and covariate (c) dependent survival prob., additive on the logit scale;

$p(t+c)$: time (t) and covariate (c) dependent capture prob., additive on the logit scale;

ψ : first-order Markov transition probabilities (time independent);

$\alpha(c)$: covariate (c) dependent assignment probabilities.

We extend the above model to allow for a semi-Markov state process. In particular, we consider each covariate dwell-time distribution to be either a (shifted) negative binomial or a geometric distribution. Thus, we initially consider four possible models. Specifying a (shifted) Poisson distribution led to significantly worse fitting models, so that we omit these results. Since it is not clear if it is reasonable to assume that the covariate process is in its stationary distribution at the initial capture occasion, we considered two different model formulations: (i) using the stationary distribution as initial distribution; (ii) estimating an additional parameter giving the probability of being in the conjunctivitis state at the initial capture, but assuming stationarity *within* state aggregates. The resulting parameter estimates obtained using the two different models were very similar to each other. In particular, for the configuration of dwell-time distributions favoured by the AIC, the probability of being in the conjunctivitis state at the initial capture was estimated as 0.07 when using (i) and as 0.05 when using (ii). In the following, we present the results obtained using the simpler model formulation (i). The `R` code used for model fitting is provided in the Web Appendix.

4.4 *Results*

Table 2 provides the estimates of the parameters of the dwell-time distributions and the corresponding AIC values for each of the models considered. The model deemed optimal specifies a shifted negative binomial dwell-time distribution for state 1 and a geometric dwell-time distribution for state 2 (though the ΔAIC values are small for each of the other

models). The fitted dwell-time distributions, for the selected hybrid Markov/semi-Markov AS model and the standard (Markov/Markov) AS model, are displayed in Figure 2. Further, we removed the dependence of the survival and recapture probabilities on time and/or state. Based on the AIC the same structural dependence is retained, though the model with time-constant survival probabilities performed only slightly worse (ΔAIC=2.941).

[Table 2 about here.]

[Figure 2 about here.]

As in previous analyses, the survival probabilities for individuals with conjunctivitis are lower than those for individuals without conjunctivitis present. The differences in the survival probability estimates obtained with the semi-Markov model and the first-order Markov model, respectively, are small (see Figure 3). The semi-Markov model predicts higher probabilities for an individual to contract conjunctivitis for short dwell times in the covariate state corresponding to the absence of conjunctivitis. This suggests that an individual that has very recently recovered from conjunctivitis has a higher probability of contracting conjunctivitis than an individual that has not suffered from the disease for a longer duration. (An alternative explanation may be that of misclassification, where individuals with conjunctivitis are misdiagnosed as healthy – we do not model possible misclassifications here). For the semi-Markov model, the stationary distribution for the states is $(0.933, 0.067)$. In other words, the stationary distribution indicates that about 6.7% of the population will have conjunctivitis at any given time (assuming the individuals observed are representative of the whole population). Assuming a first-order Markov process for both states led to a stationary distribution with 6.9% of the population with conjunctivitis.

[Figure 3 about here.]

5. Discussion

We describe a generalization of the AS model for multi-state capture-recapture-recovery data which permits the modeling of memory within the evolution of the states using semi-Markov state processes. Within the model, a parametric distribution is specified on the length of time an individual remains in each given state. Corresponding models are attractive due to the increased level of biological realism they permit with a very limited increase in the number of additional parameters, in contrast to many previous attempts of introducing memory by using higher-order Markov models. Ignoring memory present within the state process of the model can lead to biased inference, most notably with regard to the conditional state transition probabilities. Interestingly, in our simulation study, no bias was observed in the survival probabilities when assuming an incorrect first-order Markov model. However, in general the dwell-time distributions may themselves be of interest, for example, where state refers to life stage or disease status. Further, it is possible to test biological hypotheses with regard to the absence or presence of memory in the state process.

Semi-Markov multi-state capture-recapture models have previously been suggested (Choquet *et al.*, 2014). However, these models were restrictive, with the proposed model-fitting algorithm limited to a single semi-Markovian state, which could only be visited once within the study period. For the general case, Choquet *et al.* (2014) suggest tailoring the approach developed by Guédon (2003, 2005), for fitting hidden semi-Markov models using the EM algorithm, to the capture-recapture context. In contrast, our approach to inference for semi-Markov AS models builds on the approach suggested by Langrock and Zucchini (2011), where a first-order Markov process is constructed that accurately represents a given semi-Markov process. The crucial advantage of this approach is that the efficient (and simple) recursions available for first-order models become applicable also to semi-Markov models. Various similar approaches have been discussed in the literature. A good overview of these is

given in Johnson (2005), who argues that the usage of this kind of model is "almost certainly a much better practical choice for duration modeling than development and implementation of more complex and computationally expensive models with explicit modifications to handle duration probabilities" – a statement to which we agree. While in theory our approach can have computational limitations – as argued by Choquet *et al.* (2014) – those would occur only if for some of the dwell-time distributions the support would be very large, covering several hundred time units (i.e., capture occasions). This is unlikely to occur in capture-recapture scenarios, with usually only relatively short time series of capture occasions.

The semi-Markov AS model was applied to house-finch data, where an indication of memory was found within the dwell-time distribution of an individual being in the state of non-conjunctivitis. In other words, the length of time for which an individual contracts conjunctivitis may depend on when it was last infected (i.e., reinfection of conjunctivitis may be more likely immediately following their previous recovery from infection). However, the length of time an individual does suffer from conjunctivitis appears to be well modeled via a first-order Markov model (i.e., recovery from conjunctivitis is independent of the length of time an individual has suffered from conjunctivitis).

The approach described within this paper can immediately be extended to other forms of capture-recapture models, including multi-event models (Pradel, 2005) and partially observed capture-recapture-recovery models (King and McCrea, 2014), where states are not directly observed or observed with error. We note that the multi-event model has exactly the form of a standard hidden (semi-)Markov model, where no states are directly observed. In these cases, the transition process remains the same, and only the observation process needs to be modified in terms of the assignment probabilities associated with different events or partial states, conditional on being in each given state. This again emphasizes the advantage of specifying the capture-recapture(-recovery) model as a state-space (or hidden Markov)

model, separating the system and observation processes (King, 2014). Furthermore, being able to calculate the likelihood makes it straightforward to estimate model components nonparametrically via penalized likelihood (Michelot *et al.*, 2015). The extension to multiple covariate processes (for example, site and breeding status) is also conceptually straightforward, but the curse of dimensionality renders the model-fitting more computationally intensive.

The semi-Markov AS model described currently assumes that the remaining model parameters, most notably the survival probabilities, are time- and/or state-dependent. A similar semi-Markov (state-independent) model can be proposed for the survival probabilities, essentially specifying a parametric distribution on survival. However, the extension to include a state-dependent semi-Markov survival model is non-trivial if individuals can move between different states. The inclusion of additional covariate information provides a further modeling complexity of interest within such models. The development and fitting of such models is the focus of current research.

Supplementary Material

Web Appendices referenced in Sections 3 and 4 and the associated `R` computer codes are available with this paper at the Biometrics website on Wiley Online Library.

Acknowledgements

We would like to thank Evan Cooch and Paul Conn for providing the house finch data analyzed in Section 4. We would also like to thank the AE and two anonymous referees for their useful comments.

References

Brownie C, Hines JE, Nichols JD, Pollock KH, Hestbeck JB (1993), Capture-recapture studies for multiple strata including non-Markovian transitions. *Biometrics* **49**, 1173–

1187.

Catchpole EA, Freeman SN, Morgan BJT, Nash WJ (2001), Abalone I: analyzing mark-recapture-recovery data incorporating growth and delayed recovery. *Biometrics* **57**, 469–477.

Choquet R, Guédon, Y, Besnard, A, Guillemain, M, Pradel, R (2013), Estimating stop over duration in the presence of trap-effects. *Ecological Modelling* **250**, 111–118.

Choquet R, Béchet, A, Guédon, Y (2014), Applications of hidden hybrid Markov/semi-Markov models: from stopover duration to breeding success dynamics. *Ecology and Evolution* **4**, 817–826.

Conn, PB, Cooch, EG. (2009) Multistate capture-recapture analysis under imperfect state observation: an application to disease models. *Journal of Applied Ecology* **46** 486–492.

Faustino CR, Jennelle CS, Connolly V, Davis AK, Swarthout EC, Dhondt AA, Cooch EG (2004) Infection dynamics in a house finch population: seasonal variation in survival, encounter, and transmission rate. *Journal of Animal Ecology*, **73**, 651–669.

Fewster RM, Patenaude NJ (2009) Cubic splines for estimating the distribution of residence time using individual resightings data. In: Thomson, D. L., Cooch, E. G., Conroy, M. J. (eds) *Modeling Demographic Processes in Marked Populations*. Springer, Berlin, 393–415.

Gimenez O, Rossi V, Choquet R, Dehais C, Doris B, Varella H, Vila J-P, Pradel R (2007) State-space modelling of data on marked individuals. *Ecological Modelling* **206**, 431–438.

Guédon, Y (2003) Estimating hidden semi-Markov chains from discrete sequences *Journal of Computational and Graphical Statistics* **12**, 604–639.

Guédon, Y (2005) Hidden hybrid Markov/semi-Markov chains *Computational Statistics and Data Analysis* **46**, 663–688.

Hestbeck JB, Nichols JD, Malecki RA (1991) Estimates of movement and site fidelity using mark-resight data of wintering canada geese. *Ecology Letters* **72**, 523–533.

Johnson MT (2005), Capacity and complexity of HMM duration modeling techniques. *IEEE Signal Processing Letters* **12**, 407–410.

King R (2014), Statistical ecology. *Annual Review of Statistics and its Application* **1**, 401–426.

King R (2012), A review of Bayesian state-space modelling of capture-recapture-recovery data. *Interface Focus*, **2**, 190–204.

King R, Brooks SP (2003), Closed form likelihoods for Arnason-Schwarz models. *Biometrika* **90**, 435–444.

King R, Brooks SP (2002), Bayesian Model Discrimination for Multiple Strata Capture-Recapture Data. *Biometrika* **89**, 785–806.

King R, McCrea RS (2014), A generalised likelihood framework for partially observed capture-recapture-recovery models. *Statistical Methodology* **17** 30–45.

Langrock R, Zucchini W (2011), Hidden Markov models with arbitrary state dwell-time distributions. *Computational Statistics and Data Analysis* **55**, 715–724.

Langrock R, King R (2013), Maximum likelihood estimation of mark-recapture-recovery models in the presence of continuous covariates. *Annals of Applied Statistics* **7**, 1709–1732.

Langrock R, King R, Matthiopoulos J, Thomas L, Fortin D and Morales JM (2012) Flexible Hidden Markov-type Models for Animal Telemetry Data. *Ecology* **93**, 2336–42.

Lebreton JD, Nichols JD, Barker R, Pradel R, Spendelow J (2009), Modeling individual animal histories with multistate capture-recapture models. (ed Caswell H) *Advances in Ecological Research* **41**, 87–173.

Michelot T, Langrock R, Kneib T, King R (2015), Maximum penalized likelihood estimation in semiparametric capture-recapture models. *Biometrical Journal*, in press, DOI:10.1002/bimj.201400222.

Pledger SA, Efford MG, Pollock K, Collazo JA, Lyons J (2009) Stopover duration analysis with departure probability dependent on unknown time since arrival. In: Thomson, D. L., Cooch, E. G., Conroy, M. J. (eds) *Modeling Demographic Processes in Marked Populations*. Springer, Berlin, 349–363.

Pradel R (2005), Multievent: An extension of multistate capture–recapture models to uncertain states. *Biometrics* **61**, 442–447.

Pradel R (2009), The stakes of capture-recapture models with state uncertainty. In: Thomson, D. L., Cooch, E. G., Conroy, M. J. (eds) *Modeling Demographic Processes in Marked Populations*. Springer, Berlin, 781–795.

Rouan L, Choquet R, Pradel R (2009), A general framework for modeling memory in capture-recapture data. *Journal of Agricultural, Biological, and Environmental Statistics* **14**, 338–355.

Royle JA (2008) Modeling individual effects in the Cormack-Jolly-Seber model: a state-space formulation. *Biometrics* **64**, 364–370.

Schofield MR, Barker RJ (2011), Full open population capture-recapture models with individual covariates. *Journal of Agricultural, Biological, and Environmental Statistics* **16**, 253–268.

Schwarz, CJ, Schweigert, JF, Arnason, AN (1993), Estimating migration rates using tag-recovery data. *Biometrics* **49**, 177–193.

Schwarz, CJ, Seber, GAF (2000), A review of estimating animal abundance III. *Statistical Science* **14**, 427–456.

Received XXXXX 20XX. *Revised XXXXX* 20XX. *Accepted XXXXX* 20XX.

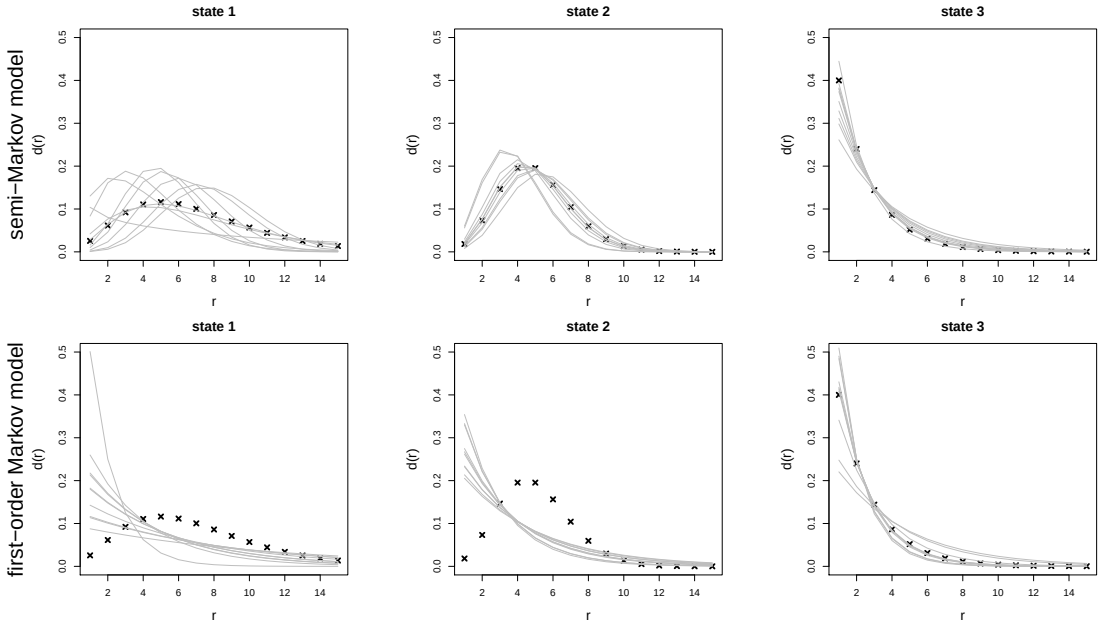

Figure 1. Simulation study: true (black crosses) state dwell-time distributions and the corresponding estimates (grey lines) obtained in the first 10 simulation runs; the top plots show the results obtained when using the semi-Markov model, and the bottom plots those when using the incorrect first-order Markov model. Note that $d(r)$ denotes the p.m.f. for the dwell-time distribution, i.e. the probability an individual stays in a given state for $r = 1, 2, \ldots$ capture events.

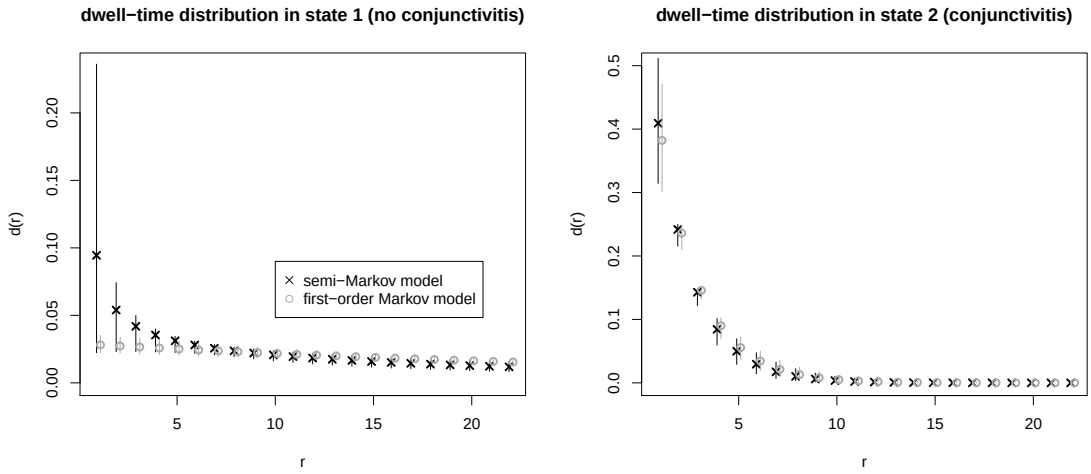

Figure 2. House finch case study: Fitted state dwell-time distributions as obtained using the semi-Markov model (with shifted negative binomial state dwell-time distribution for state 1 and a geometric state dwell-time distribution for state 2; crosses) and the first-order Markov model (circles), respectively. The vertical lines indicate 95% pointwise confidence intervals obtained using the observed Fisher information. Note that $d(r)$ denotes the p.m.f. for the dwell-time distribution, i.e., the probability an individual stays in a given state for $r = 1, 2, \ldots$ capture events.

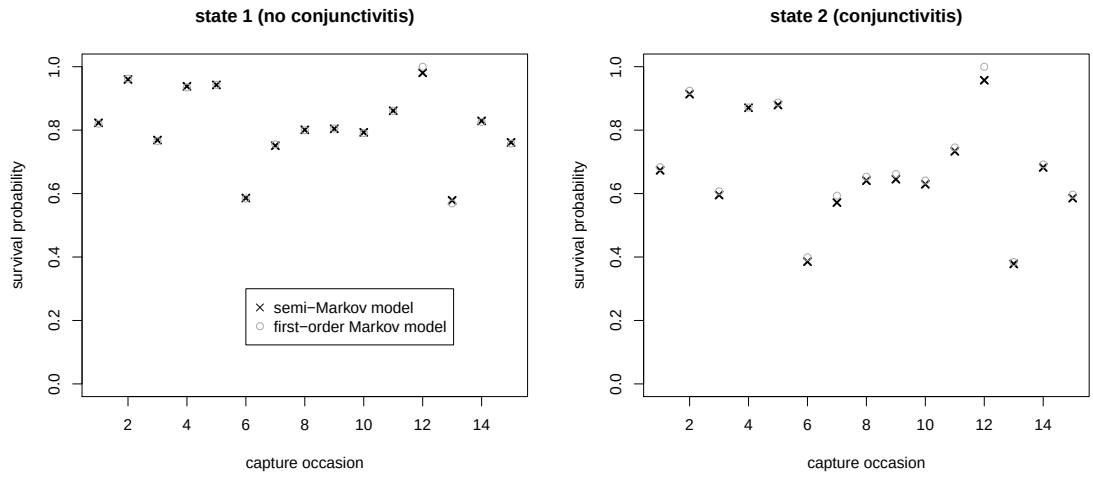

Figure 3. House finch case study: State- and time-dependent survival probabilities estimated using the semi-Markov model (crosses) and the first-order Markov model (circles), respectively.

Table 1
MRB: mean relative bias of estimates; MSD: mean standard deviation of estimates.

	semi-Markov model		first-order Markov model	
	MRB	MSD	MRB	MSD
$\hat{\psi}^*(1,2)$	-0.01	0.14	0.14	0.15
$\hat{\psi}^*(2,1)$	0.01	0.10	0.05	0.10
$\hat{\psi}^*(3,1)$	0.01	0.22	-0.16	0.27
$\hat{\phi}(1)$	0.00	0.03	0.00	0.04
$\hat{\phi}(2)$	0.00	0.05	0.00	0.05
$\hat{\phi}(3)$	-0.01	0.08	-0.01	0.08
$\hat{p}(1)$	0.03	0.04	0.09	0.07
$\hat{p}(2)$	0.05	0.03	0.05	0.05
$\hat{p}(3)$	0.11	0.19	0.15	0.20
$\hat{\lambda}$	0.01	0.02	0.00	0.02

Table 2

Model selection for different dwell-time distributions for the house finch data using the AIC statistic, where state 1 corresponds to the absence of conjunctivitis and state 2 to the presence of conjunctivitis. The MLEs of the parameters for the dwell-time distributions are provided in brackets. The remaining model parameters are of the following form: $\phi(t+k)/p(t+k)/\alpha(k)$.

Dwell-time distributions			
State 1	State 2	AIC	ΔAIC
geom(0.028)	geom(0.382)	5539.211	0.626
sNB(0.581,0.017)	geom(0.409)	**5538.585**	**0.0**
geom(0.029)	sNB(0.612,0.287)	5540.508	1.923
sNB(0.536,0.016)	sNB(0.463,0.265)	5539.144	0.559